FoundPose: Unseen Object Pose Estimation with Foundation Features (Supplementary Material)

Evin Pınar Örnek[1] Yann Labbé[2] Bugra Tekin[2] Lingni Ma[2]
Cem Keskin[2] Christian Forster[2] Tomas Hodan[2]

[1]Technical University of Munich [2]Meta Reality Labs

This supplement provides additional qualitative results of FoundPose, the proposed method for 6D pose estimation of unseen objects, and describes the supplementary video.

1 Additional qualitative evaluation

The following pages show qualitative results on the seven core BOP datasets: LM-O [1], T-LESS [5], TUD-L [6], IC-BIN [2], ITODD [3], HB [7], and YCB-V [8]. As in Fig. 4 of the paper, each example shows the query image crop with the CNOS mask in white (top left), retrieved templates (middle row), and matched patch descriptors of the crop and the template that led to a pose with the highest number of inlier correspondences (bottom row). The contour of the 3D object model in the ground-truth pose is shown in red, the coarse pose in blue, and the refined pose in green (top right). The ground truth is not publicly available for ITODD and HB and, therefore, not shown. To make the bottom row less cluttered, we show only up to 100 matches with the smallest cyclical distance of patch descriptors [4]. The colors of patch descriptors are given by the top three PCA components calculated from patch descriptors extracted from all templates of the given object.

We show nine examples per dataset. The first six are examples where the refined pose is close to the ground-truth pose, while the last three are examples where the final pose is less accurate due to challenges such as imperfect segmentation masks or heavy occlusion.

2 Supplementary video

The attached video shows the same qualitative results as on the following pages, with each result followed by an animation of the featuremetric pose refinement process. The first part of the video shows six examples per dataset with more accurate estimates, while the second part (starting at 7:55) shows three examples per dataset with less accurate estimates. For each refinement iteration, the animation shows the model contour in the current pose estimate in green and the model contour in the ground-truth pose in red. The points at which we calculate the featuremetric error (magenta represents low while cyan represents high per-point error) are also shown. The sum of the per-point errors, which is shown in a text overlay, defines the error that is minimized in the process (see Sec. 3.5 of the paper). Note that the refinement process was run for up to 30 iterations to obtain the final pose, while the animation shows only up to 15 iterations.

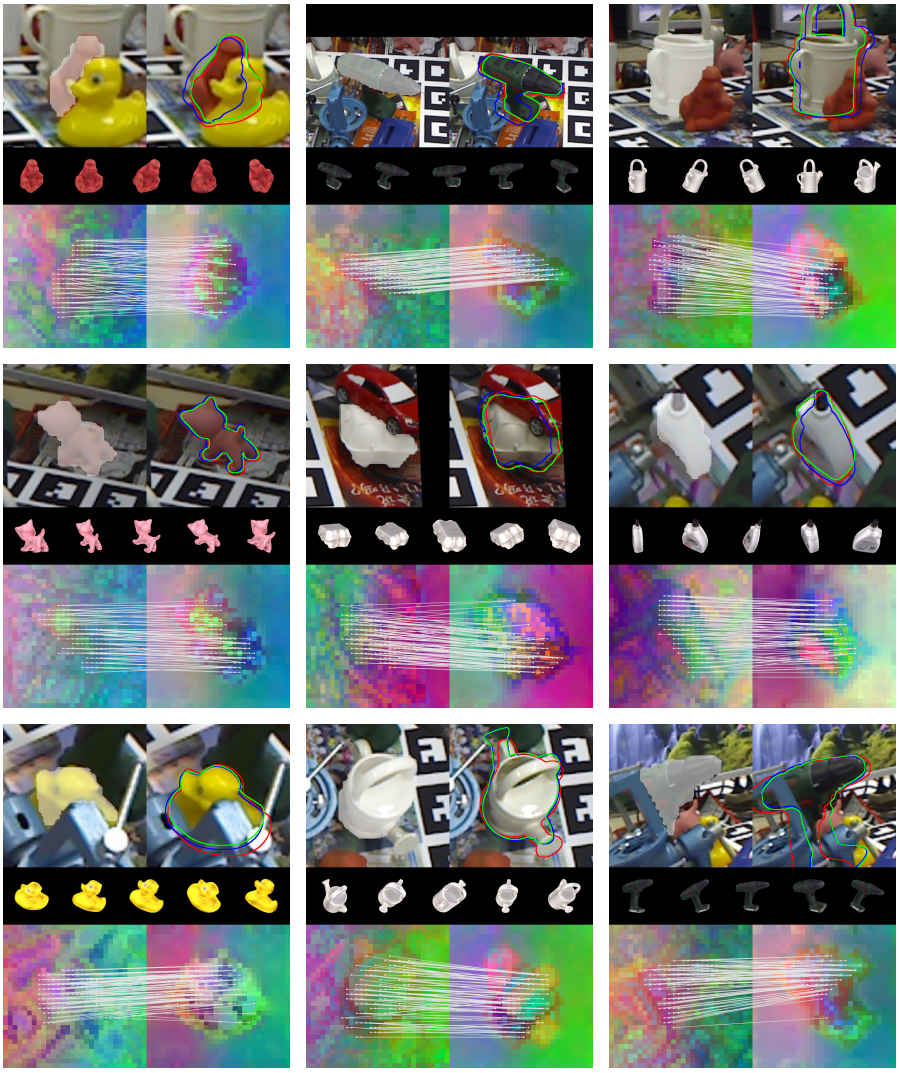

Fig. 1: Example FoundPose results on the LM-O dataset [1].

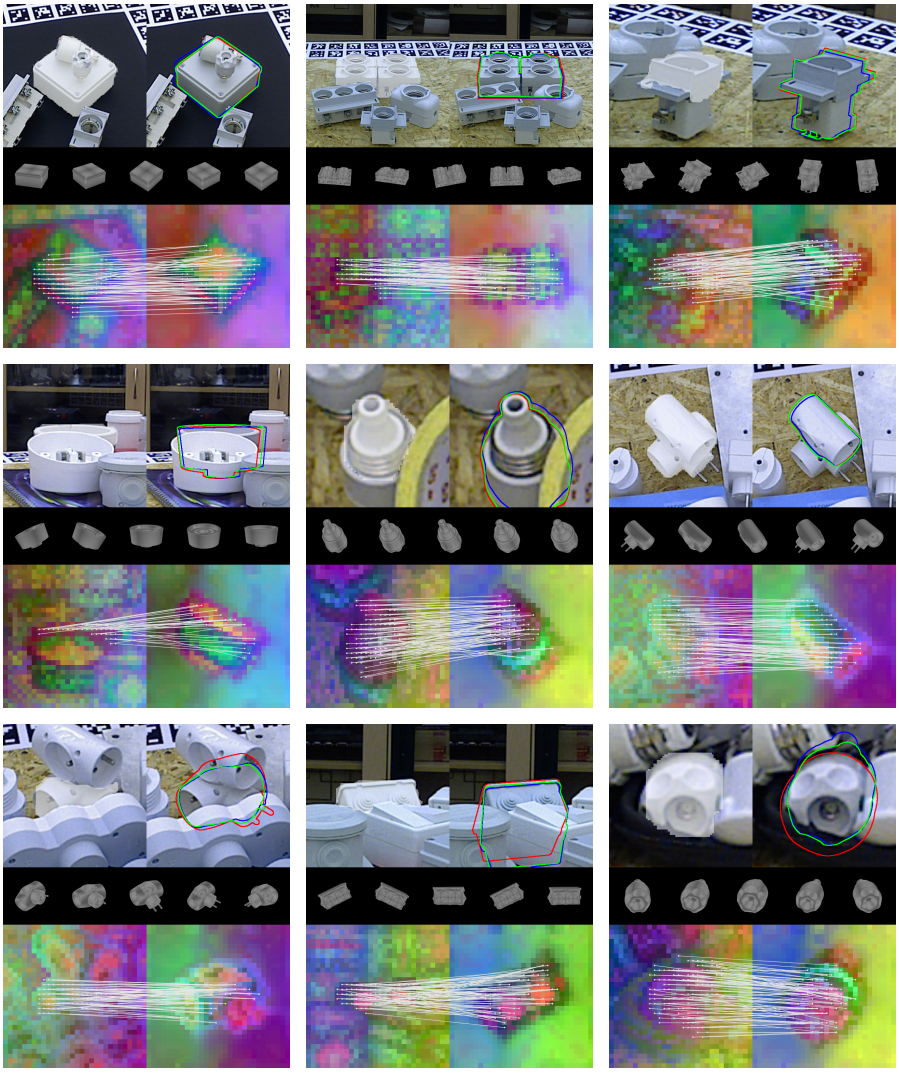

Fig. 2: Example FoundPose results on the T-LESS dataset [5].

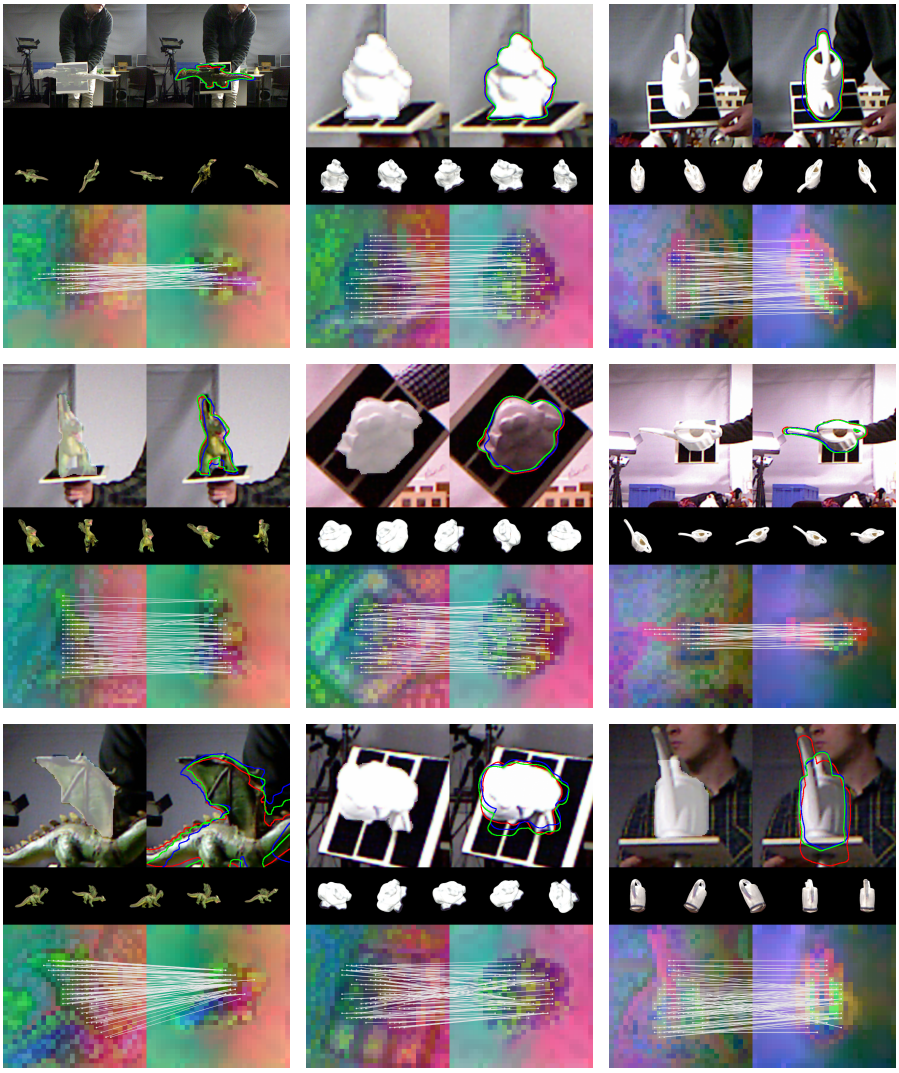

Fig. 3: Example FoundPose results on the TUD-L dataset [6].

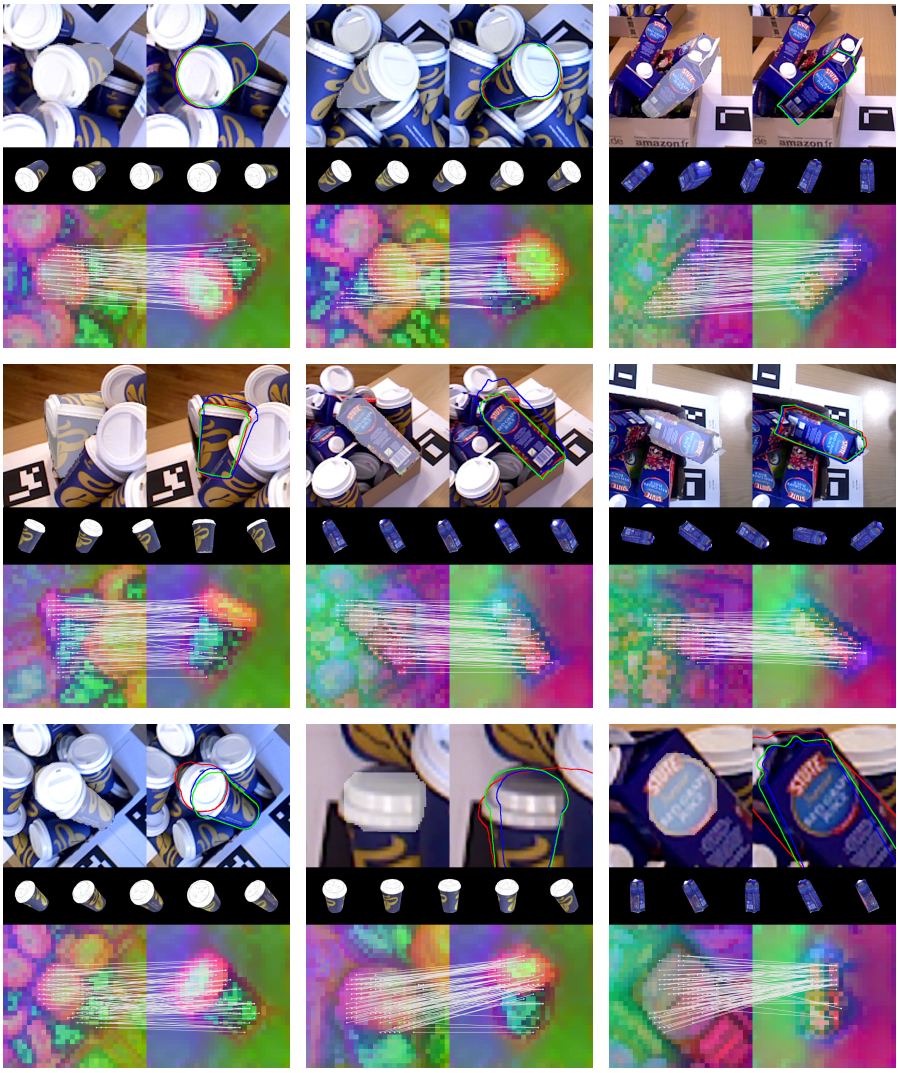

Fig. 4: Example FoundPose results on the IC-BIN dataset [2].

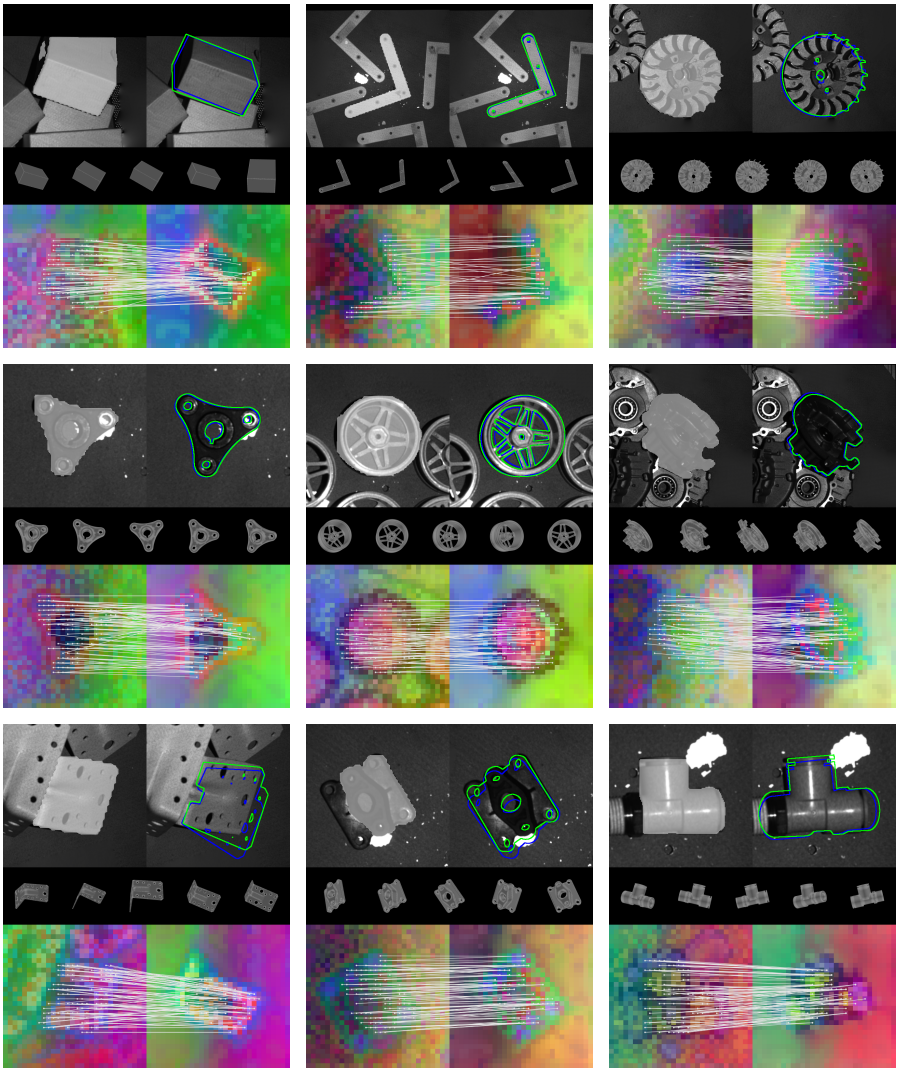

Fig. 5: Example FoundPose results on the ITODD dataset [3].

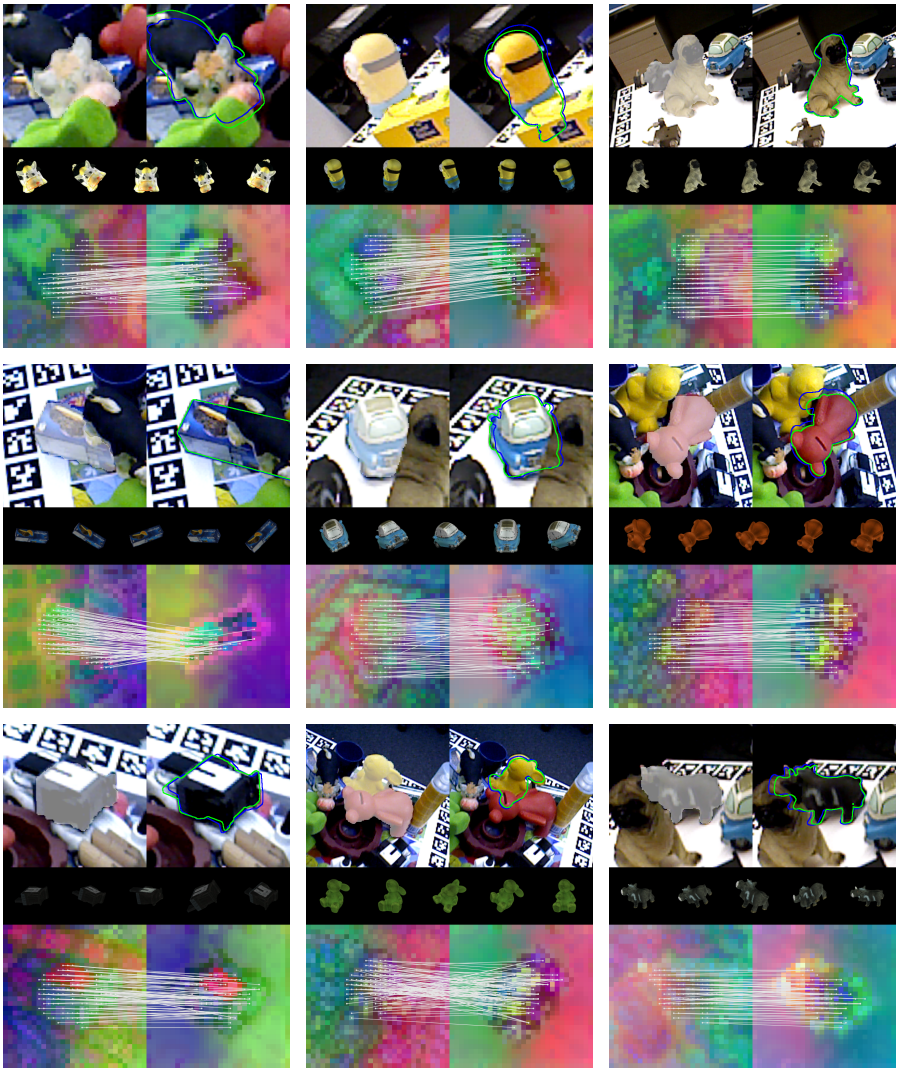

Fig. 6: Example FoundPose results on the HB dataset [7].

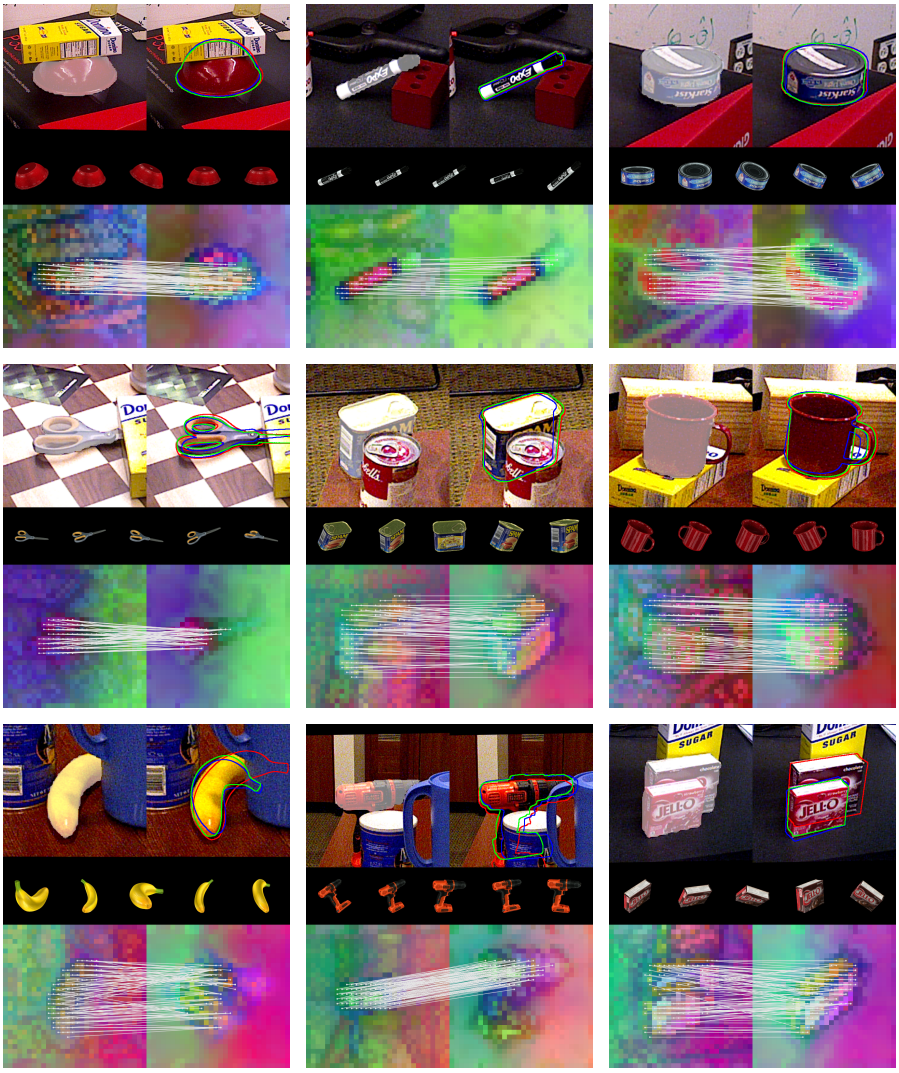

Fig. 7: Example FoundPose results on the YCB-V dataset [8].

References

1. Brachmann, E., Krull, A., Michel, F., Gumhold, S., Shotton, J., Rother, C.: Learning 6D object pose estimation using 3D object coordinates. ECCV (2014) 1, 2
2. Doumanoglou, A., Kouskouridas, R., Malassiotis, S., Kim, T.K.: Recovering 6D object pose and predicting next-best-view in the crowd. CVPR (2016) 1, 5
3. Drost, B., Ulrich, M., Bergmann, P., Hartinger, P., Steger, C.: Introducing MVTec ITODD – A dataset for 3D object recognition in industry. ICCVW (2017) 1, 6
4. Goodwin, W., Vaze, S., Havoutis, I., Posner, I.: Zero-shot category-level object pose estimation. ECCV (2022) 1
5. Hodan, T., Haluza, P., Obdrzalek, S., Matas, J., Lourakis, M., Zabulis, X.: T-LESS: An RGB-D dataset for 6D pose estimation of texture-less objects. WACV (2017) 1, 3
6. Hodan, T., Michel, F., Brachmann, E., Kehl, W., Glent Buch, A., Kraft, D., Drost, B., Vidal, J., Ihrke, S., Zabulis, X., Sahin, C., Manhardt, F., Tombari, F., Kim, T.K., Matas, J., Rother, C.: BOP: Benchmark for 6D object pose estimation. ECCV (2018) 1, 4
7. Kaskman, R., Zakharov, S., Shugurov, I., Ilic, S.: HomebrewedDB: RGB-D dataset for 6D pose estimation of 3D objects. ICCVW (2019) 1, 7
8. Xiang, Y., Schmidt, T., Narayanan, V., Fox, D.: PoseCNN: A convolutional neural network for 6D object pose estimation in cluttered scenes. RSS (2018) 1, 8